

# Weekly Report

Pingping Shang

2013.10.28~2013.11.3

## 本周工作

### 1. 最初想法：

根据 CCA 得到最大相关系数，然后利用这些相关系数得到 block 的聚类或是用于 block 重排序；通过 CCA 得到最大负相关系数，得到互斥的人群或商品；由最大相关系数对应的变量系数得到对相关关系贡献最大的商品

调研加试验结论：

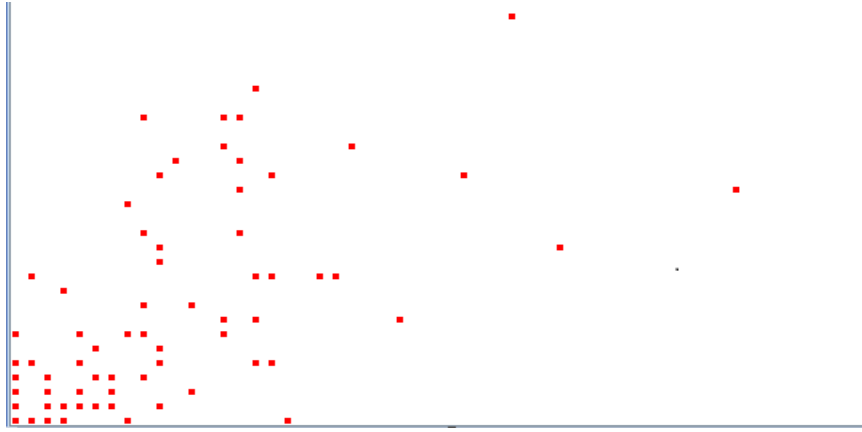
- 1) 经过查阅资料，CCA 得到的各变量前的系数只是系数，不代表权重，不代表重要性，所以之前讨论的通过系数得到最影响人群行为的商品是行不通的。
  - 2) 由 CCA 的计算过程得出 CCA 计算得到的相关系数都是非负的。而且计算得到的第一最大相关系数都是较大的，大部分在 0.7 以上。作为 MDS 的距离矩阵得出的结果很不理想。
2. 1) 通过 CCA 计算得到人群的相关系数，但发现得到的相关系数都很大，大部分都在 0.7 以上，将 block 的相关系数作为相似性矩阵得到 MDS 投影结果，并不能将 block 分开，而且有时 MDS 收敛不到一个正确的结果。
- 2) 我们还试验了 block 间的皮尔森相关系数：将一个 block 作为一个向量，计算任两个 block 之间的皮尔森相关系数.得到的结果同样是都很大。

$$r = \frac{1}{n-1} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{s_X} \right) \left( \frac{Y_i - \bar{Y}}{s_Y} \right)$$

皮尔森相关系数定义：

为了直观上对相关系数的大小有个理解，我将两个肉眼看上去最不相似的两个 block 的原始数据投在画布上（将第一个和第二个 block 分别看作 x 变量和 y 变量），还是可以明显观察到有一定线性关系的，计算得到的这两个 block 的皮尔森相关系数为 0.59。





## 下周工作

实现下之前和解聪讨论的矩阵变换操作，但是矩阵变换时根据哪种相似度进行变换还需进一步讨论。